A NEW APPROACH TO THE DISTRIBUTION OF PRIME NUMBERS

YASUO NISHII

Abstract. The aim of the thesis is to determine the distribution of prime numbers by a method induced from the nature of prime numbers which take infinite procudures to be identified when primes are arbitrarily large. The first section is devoted to the introduction. In the second section we mention that the probability is induced when procedures of verifying an assumption are infinite. And we state some results of the probability theory in the third section. At the last section we prove some theorems which determine the distribution of prime numbers.

Contents

Date: April 1, 2014.

1. Introduction

The distribution of prime numbers is mainly studied in the two fields, that is, in complex analysis and in number theory and many results are obtained from these fields. The way in the thesis is, however, neither. Instead, we use the method which is close to algorithm theory since we use the algorithm which procedures are considerably large. Actually, it takes infinite steps to identify the prime numbers when primes are arbitrarily large. This nature of prime numbers is related to the probability theory as we state in the second section. To use this method, we must state some fundamental results of the probability theory in the third section. Finally we can prove some theorems about the distribution of prime numbers in the last section.

2. The relation between procedures and probabilities

In this section, f represents a function which has the domain of natural numbers and the range of real numbers.

In order to eliminate the unclearness of the foundation of the theory, we need to decide the condition where the probability is induced. So we state the following theorems.

Theorem 1. *If we cannot verify $f(n) = a$ or $f(n) \neq a$ in finite procedures, the probability of $f(n) = a$ is p such that $0 < p < 1$.*

Proof. If the probabilty of $f(n) = a$ is 0 or 1, we can decide $f(n) = a$ or $f(n) \neq a$ in finite steps because whether $f(n) = a$ is decided. This is contradiction. □

Theorem 2. *If the probability of $f(n) = a$ is p such that $0 < p < 1$, we cannot verify $f(n) = a$ or $f(n) \neq a$ in finite procedures.*

Proof. This is the converse of Theorem 1. Suppose we can determine $f(n) = a$ or not in finite procedures, then $p = 0$ or $p = 1$. This is contradiction. □

Corollary 1. *If it takes infinite procedures to verify whether a certain assumption holds, the probability of the assumption is p such that $0 < p < 1$.*

Proof. Deduction from Theorem 1. □

Corollary 2. *If the probability of a certain assumption is p such that $0 < p < 1$, we cannot verify whether the assumption holds.*

Proof. Deduction from Theorem 2. □

We define that f has infinite procedures when the following conditions are satisfied.

Definition 1. *We define that f has infinte procedures when we can choose the subsequence $\{n_i\}$ from $\mathbb{N}$ such that the procedures of identifying $f(n_i)$ goes to infinity when n_i goes to infinity.*

We can consider that $f(n)$ has the probability distribution for each n when f has infinite procedures.

3. Fundamental results of the probability theory

Throughout this section, $P(A)$ stands for the the probability of A and $E(X)$ for the expected value of X. f represents a function which takes random variables as its values with the domain of natural numbers and satisfies $P(f(n)=0)+P(f(n)=1)=1$ for each n. The section is devoted to the proof which shows the relation between the expectated value $\sum E(f(n))$ and the estimate of the error term $|\sum f(n) - \sum E(f(n))|$. First we state the next proposition.

Proposition 1. *Let $0 \leq p_1, \cdots, p_k \leq 1$ and $p_1 + \cdots + p_k = kp = m$, then $\sum_{(i_1,\cdots,i_q)\subset(1,\cdots,k)}(\prod_{l=1}^{q} p_{i_l} \prod_{l=q+1}^{k}(1-p_{i_l})) \leq \binom{k}{q} p^q(1-p)^{k-q}$ for $0 \leq q \leq k$ and $q < m - A\sqrt{m}$ or $q > m + A\sqrt{m}$ where m is adequately large and A is some constant value.*

Proof. We can prove the proposition by applying Lagrange's method of indeterminate coefficients to it. Let

$$h(p_1, \cdots, p_k) = \sum_{(i_1,\cdots,i_q)\subset(1,\cdots,k)} (\prod_{l=1}^{q} p_{i_l} \prod_{l=q+1}^{k} (1-p_{i_l})). \tag{1}$$

Then we can say that it takes relative extremum at the point $(p_1, \cdots, p_k) = (p, \cdots, p)$ by the patial differentiation of the expression $h - \lambda(m - (p_1 + \cdots + p_k))$ by $\lambda, p_1, \cdots, p_k$. From now on, we assume $s = (s_1, \cdots, s_k)$ is maximal point of h on condition $0 \leq s_1, \cdots, s_k \leq 1$ and $s_1 + \cdots + s_k = kp = m$ and we suppose $0 \leq q \leq k$ and $q < m - A\sqrt{m}$ or $q > m + A\sqrt{m}$ for some $A > 1$. We define $M_h(m) = h(s)$. Next we suppose s_{n_1} and s_{n_2} are not the same values for some n_1 and n_2 such that $1 \leq n_1 < n_2 \leq k$. Then the partial differentiation of $h - \lambda(m - (p_1 + \cdots + p_k))$ by p_{n_1} at s induces $\partial h(s)/\partial p_{n_1} + \lambda = 0$ and the partial differentiation by p_{n_2} at s induces $\partial h(s)/\partial p_{n_2} + \lambda = 0$, so $\partial h(s)/\partial p_{n_1} = \partial h(s)/\partial p_{n_2}$. Since h is linear and symmetric for each p_i, h is represented as $h(p_1, \cdots, p_k) = c_2 p_{n_1} p_{n_2} + c_1(p_{n_1} + p_{n_2}) + c_0$ for some function c_0, c_1 and c_2 where they do not contain p_{n_1} and p_{n_2} as variables. We can say $c_2 s_{n_1} + c_1 = c_2 s_{n_2} + c_1$ since $\partial h(s)/\partial p_{n_1} = \partial h(s)/\partial p_{n_2}$. Then $c_2 = 0$ because $s_{n_1} \neq s_{n_2}$. So $\partial^2 h(s)/\partial p_{n_1} \partial p_{n_2} = 0$.

Suppose all s_i and s_j such that $s_i \neq s_j$ are $s_i + s_j \leq p$, then $s_i < p/2$ for all s_i so $m < pk/2$. But $pk = m$. This is contradiction. Next suppose all s_i and s_j satisfy $s_i + s_j \geq 1 + p$, then $m > (p+1)k/2 = m/2 + k/2$, that is, $m > k$. This is also contradiction. If suppose we cannot choose s_{n_1} and s_{n_2} such that $1 + p > s_{n_1} + s_{n_2} > p$, then we can choose s_{a_1}, s_{a_2}, s_{b_1} and s_{b_2} which satisfy $s_{a_1} + s_{a_2} \leq p$ and $s_{b_1} + s_{b_2} \geq 1 + p$ where $s_{a_1} \neq s_{a_2}$ and $s_{b_1} \neq s_{b_2}$ because some s_i and s_j must satisfy $s_i + s_j \geq 1 + p$ or $s_i + s_j \leq p$ by the above statements. Then $0 \leq \min(s_{a_1}, s_{a_2}) < p/2 < (1+p)/2 < \max(s_{b_1}, s_{b_2}) \leq 1$. So $p \leq (1+p)/2 < \min(s_{a_1}, s_{a_2}) + \max(s_{b_1}, s_{b_2}) < 1 + p/2 \leq 1 + p$. Then $s_{n_1} = \min(s_{a_1}, s_{a_2})$ and $s_{n_2} = \max(s_{b_1}, s_{b_2})$ induce $p < s_{n_1} + p_{n_2} < 1 + p$. But this is contradiction. Therefore $p < s_{n_1} + s_{n_2} < p + 1$ and $h(s_1, \cdots, s_k) = h(s'_1, \cdots, s'_k)$ such that $s'_{n_1} = s_{n_1} + s_{n_2} - p$, $s'_{n_2} = p$ and $s'_i = s_i$ where $1 \leq i \neq n_1, n_2 \leq k$. This fact can be induced by the equation $h = c_1(s_{n_1} + s_{n_2}) + c_0 = c_1(s'_{n_1} + s'_{n_2}) + c_0$. Since we assumed $h(s_1, \cdots, s_k) = M_h(m)$,

we can say $h(s_1, \cdots, s_k) = h(s'_1, \cdots, s'_k)$. Repeating this operation at other points, we can conclude $h(s_1, \cdots, s_k) = h(p, \cdots, p)$.

We must verify $h(p, \cdots, p) = M_h(m)$. Suppose $2 \leq q \leq k-2$ and $q < m - A\sqrt{m}$ or $m + A\sqrt{m} < q$ for some $A > 1$. Substituting $p_k = m - p_1 - \cdots - p_{k-1}$ into the expression $h(p_1, \cdots, p_k)$ and differentiating two times at $(p_1, \cdots, p_{k-1}) = (p, \cdots, p)$ by some p_i such that $1 \leq i \leq k-1$, we obtain $-2\binom{k-2}{q-2}p^{q-2}(1-p)^{k-q} + 4\binom{k-2}{q-1}p^{q-1}(1-p)^{k-q-1} - 2\binom{k-2}{q}p^q(1-p)^{k-q-2}$. Comparing the coefficients, it is decided whether the value is the maximum or not by the expression

$$-\frac{q-1}{k-q}\frac{1-p}{p} + 2 - \frac{k-q-1}{q}\frac{p}{1-p} \tag{2}$$

being negative or positive. We suppose $p \neq 0, 1$ since we do not have to considr the case when $p = 0, 1$ because the proposition is satisfied when $p = 0, 1$ irrespective of $q < m - A\sqrt{m}$ or $q > m + A\sqrt{m}$. Solving the equation (2) $= 0$ for q, we can obtain

$$q = m - \frac{1-a}{2(1+a)} \pm \sqrt{\frac{a}{1+a}m + \left(\frac{1-a}{2(1+a)}\right)^2}. \tag{3}$$

where $a = (1-p)/p > 0$, $1 > a/(1+a) > 0$ and $1/2 > (1-a)/2(1+a) > -1/2$. Let $q_\pm$ be the above solutions. The solutions $q_\pm$ are the dividing points of deciding the muximum or not. Suppose $2 \leq q < q_- \leq k-1$, then the value of (2) is minus and we obtain the muximum. Since $m - A\sqrt{m} < q_\pm < m + A\sqrt{m}$ for adequately large m and for some $A > 1$, $M_h(m) = h(p, \cdots, p)$ $2 \leq q \leq k-2$ and $q < m + A\sqrt{m}$ or $m + A\sqrt{m} < q$.

If $q = 1$, then it is decided whether $h(p, \cdots, p) = M_h(m)$ or not by the expression

$$2 - \frac{k-q-1}{q}\frac{p}{1-p} \tag{4}$$

being negative or positive. We supposed $q = 1$, so the above expression is $2 - p(k-2)/(1-p) = (2-m)/(1-p)$. Therefore it is minus since m is adequately large. Then $M_h(m) = h(p, \cdots, p)$ and $q = 1 < m - A\sqrt{m}$.

If $q = k-1$, the expression

$$-\frac{q-1}{k-q}\frac{1-p}{p} + 2 \tag{5}$$

matters. If $k > m+2$ then it is minus. We supposed $k - 1 = q > m + A\sqrt{m}$ or $k - 1 = q < m - A\sqrt{m}$, but the latter possibility is eliminated since $q = k - 1 > kp - A\sqrt{m} = m - A\sqrt{m}$. Therefore $k > m + 2$ is induced from the former inequality because m is adequately large and $A > 1$ is some constant.

So $M_h(m) = h(p, \cdots, p)$ when $1 \leq q \leq k-1$ and $q < m - A\sqrt{m}$ or $q > m + A\sqrt{m}$.

It is clear when $q = 0$ or k. So the proposition is satisfied.

□

Proposition 2. *Let $E(f(i)) = p$ for $1 \le i \le n$, $m = \sum_{i=1}^{n} E(f(i))$ and m, $n - m$ are adequately large. Then $|\sum_{i=1}^{n} f(i) - \sum_{i=1}^{n} E(f(i))| < M(m)\sqrt{m}$ where $M(x)$ is an arbitrary function that satisfies $\lim_{x\to+\infty} M(x) = +\infty$.*

Proof. By de Moivre-Laplace's theorem, the distribution of the $S_n = \sum f(i)$ goes to Gaussian distribution when $np(1-p)$ goes to infinity. Precisely, $P(a < (S_n - np)/\sqrt{np(1-p)} < b) = 1/\sqrt{2\pi}\int_a^b e^{-u^2/2}du$ for finite a, b. Since $M(x)$ goes to infinity, $P(-M(x)/\sqrt{1-p} < (S_n - np)/\sqrt{np(1-p)} < M(x)/\sqrt{1-p}) \to 1$ when $x \to \infty$ because we can take arbitrary $0 < \epsilon < 1$ which satisfies $\int_{a(\epsilon)}^{b(\epsilon)} e^{-u^2/2}du = 1-\epsilon < \int_{-M(x)/\sqrt{1-p}}^{M(x)/\sqrt{1-p}} e^{-u^2/2}du$. So $\lim P(|S_n - np| < M(m)\sqrt{m}) = 1$. This fact indicates the proposition since $S_n = \sum f(i)$ and $np = m = \sum E(f(i))$. $\square$

Combining Proposition 1 and 2, we can state the next theorem.

Theorem 3. *Let $m = \sum_{i=1}^{n} E(f(i))$ and m, $n - m$ are adequately large. Then $|\sum_{i=1}^{k} f(i) - \sum_{i=1}^{k} E(f(i))| < M(m)\sqrt{m}$ where $M(x)$ is an arbitrary function that satisfies $\lim_{x\to+\infty} M(x) = +\infty$.*

Proof. Let $P_f = \lim_{m\to\infty} P(m - M(m)\sqrt{m} < \sum f(i) < m + M(m)\sqrt{m})$. Proposition 2 means $P_f = 1$ if $P(f(i) = 1) = p$ for $1 \le i \le n$. Then Proposition 1 indicates $\inf\{P_f | 1 \le i \le n, \sum P(f(i) = 1) = m\} = 1$ since $\lim M(m) > A$. Then the theorem follows. $\square$

4. The distribution of prime numbers

We define the function *prime* as $prime(n) = 0$ if n is not a prime number and as $prime(n) = 1$ if n is a prime number where the domain of *prime* is natural numbers.

Lemma 1. *prime has infinite procedures.*

Proof. If n is a prime number and n goes to infinity, we have to implement infinite calculations in order to decide whether n is a prime number or not because we must verify whether n is divided by infinite prime numbers below n. This means the lemma. $\square$

We can consider *prime* has the probability distribution for each n according to the last part in the second section.

Let $\pi(n) = \sum_{x=2}^{n} prime(x)$ and $Li(n) = \int_2^n 1/\log(x)dx$.

Lemma 2. *All procedures of verifying whether $\pi(n_1) - \pi(n_0) > \int_{n_0}^{n_1} 1/\log n$ become infinite when $n_0, n_1, n_1 - n_0 \to +\infty$.*

Proof. The inequality in the lemma is determined by adequately large part of the prime numbers below n_0 from sieve of Eratosthenes and from the fact that $\pi(n) - Li(n)$ changes the signs infinitely. See [1]. Since that part is determined by the smaller part of prime numbers and their calculations, the procedures of obtaining the former are larger than that of the latter. So if n_0, n_1 and $n_1 - n_0$ go to infinity, steps of procedures also increase to infinity. Then the lemma follows. $\square$

Theorem 4. *$|\pi(n) - Li(n)| < 2M(m)\sqrt{m}$ for $n > N$ where $m = E(\pi(n))$, N is adequately large and $M(n)$ is the function in the third section.*

Proof. Suppose $|\pi(n) - Li(n)| \geq 2M(m)\sqrt{m}$ for arbitrarily large n. Since $|\pi(n) - E(\pi(n))| + |E(\pi(n)) - Li(n)| \geq |\pi(n) - Li(n)| \geq 2M(m)\sqrt{m}$ and $|\pi(n) - E(\pi(n))| < M(m)\sqrt{m}$ is induced from the result from Theorem 3 where m, $n - m$ are adequately large, we can deduce $|E(\pi(n)) - Li(n)| > M(m)\sqrt{m}$. So we can obtain $Li(n) < m - M(m)\sqrt{m}$ or $Li(n) > m + M(m)\sqrt{m}$ for arbitrarily large n and can also obtain $m - M(m)\sqrt{m} < \pi(n) < m + M(m)\sqrt{m}$ from the proof of Theorem 3. Then we can choose as many and as large n_k as possible on condition that the probability $Li(n_k) < \pi(n_k)$ is 0 or 1 for each n_k. Therefore we can choose more than three n_{q_1}, n_{q_2}, n_{q_3} from adequately large n_k and then we chose again k_1 and k_2 among q_1, q_2, q_3 and they satisfy $P(Li(n_{k_1}) - Li(n_{k_2}) < \pi(n_{k_1}) - \pi(n_{k_2})) = 0$ or 1 if we can choose k_1 and k_2 properly. But Lemma 2 indicates that if $n_0, n_1, n_1 - n_0 \to +\infty$ then that the probability p of $\pi(n_1) - \pi(n_0) < \int_{n_0}^{n_1} 1/\log(x)dx$ is $0 < p < 1$. If we substitute k_1 and k_2 for n_0 and n_1 each, this substitution leads to contradiction since p has to be 0 or 1. Then the theorem follows. □

Theorem 5. $|\pi(n) - Li(n)| < Q(n)\sqrt{Li(n)}$ *for* $n > N$ *where* N *is adequately large and* $Q(x)$ *is an arbitrary function that satisfies* $\lim_{x\to+\infty} Q(x) = +\infty$.

Proof. Since $|E(\pi(n)) - Li(n)| \leq |E(\pi(n)) - \pi(n)| + |\pi(n) - Li(n)| < 3M(m)\sqrt{m}$ is induced from Theorem 3 and 4, $m - 3M(m)\sqrt{m} < Li(n) < m + 3M(m)\sqrt{m}$. We can conclude $m < 2m - 6M(m)\sqrt{m} < 2Li(n)$ if $M(m) < \sqrt{m}/6$. So $|\pi(n) - Li(n)| < 2M(m)\sqrt{m} < 2M(2Li(n))\sqrt{2Li(n)}$. If we choose $Q(n) \geq 2\sqrt{2}M(2Li(n))$, then the theorem follows since $M(x)$ is arbitrary function which satisfies $M(x) < \sqrt{x}/6$ and $\lim_{x\to+\infty} M(x) = +\infty$. □

References

[1] A. E. Ingham, *The distribution of prime numbers*, Cambridge University Press, 1932.